Supplement to ”Nonlinear dimension reduction for regression with nearest neighbors”

Quentin Paris

Supplement to Nonlinear dimension reduction for regression with nearest neighbors

Quentin PARIS

IRMAR, ENS Cachan Bretagne, CNRS, UEB
Campus de Ker Lann
Avenue Robert Schuman, 35170 Bruz, France
quentin.paris@bretagne.ens-cachan.fr

1 Proof of Lemma 8.3

Lemma 1.1 *Suppose assumption* $(\mathbf{A3})$ *holds and* $d \geq 3$. *Then, for all* $h \in \mathscr{H}_d$, *we have*

$$\mathbb{E}\|X^h_{(1)}(X^h) - X^h\|^2 \leq \frac{16R^2}{n^{2/d}} \left(\frac{d}{2} - 1\right)^{4/d}.$$

Proof – Fix $h \in \mathscr{H}_d$ and denote μ_h the distribution of X^h. Since $\dim S(h) \leq d$, one may consider that X^h is $\mathbb{R}^d$-valued and that μ_h is of support in $\mathbb{R}^d$. Hence, for all $\varepsilon > 0$, we have

$$\begin{aligned}
\mathbb{P}\left(\|X^h_{(1)}(X^h) - X^h\| > \varepsilon\right) &= \mathbb{E}\Big[\mathbb{P}\left(\|X^h_{(1)}(X^h) - X^h\| > \varepsilon \big| X\right)\Big] \\
&= \mathbb{E}\Big[\mathbb{P}\left(\|X^h_1 - X^h\| > \varepsilon \big| X\right)^n\Big] \\
&= \mathbb{E}\Big[\left(1 - \mu_h\left(B_d\left(X^h, \varepsilon\right)\right)\right)^n\Big] \\
&= \int_{\|u\|_d \leq R} \left(1 - \mu_h\left(B_d(u, \varepsilon)\right)\right)^n \mu_h(\mathrm{d}u),
\end{aligned}$$

since, under assumption $(\mathbf{A3})$, the variable X^h is $B_d(0,R)$-valued. Now, for all $\varepsilon > 0$, the ε-covering number $N(\varepsilon)$ of $B_d(0,R)$ satisfies

$$N(\varepsilon) \leq \left(\frac{4R}{\varepsilon}\right)^d$$

(see *e.g.* Proposition 5 in Cucker and Smale, 2001). Thus, given $\varepsilon > 0$, one may find a finite collection of Euclidian balls $B_1, \dots, B_{N(\varepsilon)}$ of radius at most ε in $\mathbb{R}^d$

such that

$$B_d(0,R) \subset \bigcup_{i=1}^{N(\varepsilon)} B_i.$$

We can notice that for all $i \in \{1, \dots, N(\varepsilon)\}$, we have $u \in B_i \Rightarrow B_i \subset B_d(u, \varepsilon)$. Then

$$\begin{aligned}
\mathbb{P}\left(\|X^h_{(1)}(X^h) - X^h\| > \varepsilon\right) &= \int_{\|u\|_d \le R} \left(1 - \mu_h\left(B_d(u,\varepsilon)\right)\right)^n \mu_h(\mathrm{d}u) \\
&\le \sum_{i=1}^{N(\varepsilon)} \int_{B_i} \left(1 - \mu_h(B_i)\right)^n \mu_h(\mathrm{d}u) \\
&= \sum_{i=1}^{N(\varepsilon)} \mu_h(B_i)\left(1 - \mu_h(B_i)\right)^n.
\end{aligned}$$

Then, since for all $t \in [0,1]$ we have $t(1-t)^n \le \frac{1}{n}$, it follows that for all $\varepsilon > 0$

$$\mathbb{P}\left(\|X^h_{(1)}(X^h) - X^h\| > \varepsilon\right) \le \frac{N(\varepsilon)}{n} \le \frac{1}{n}\left(\frac{4R}{\varepsilon}\right)^d.$$

Now write $C_n := \frac{(4R)^d}{n}$. Using the fact that $d \ge 3$, we have

$$\begin{aligned}
\mathbb{E}\|X^h_{(1)}(X^h) - X^h\|^2 &= \int_0^{+\infty} \mathbb{P}\left(\|X^h_{(1)}(X^h) - X^h\| > \sqrt{\varepsilon}\right) \mathrm{d}\varepsilon \\
&\le \int_0^{+\infty} \min\left(1, \frac{C_n}{\varepsilon^{d/2}}\right) \mathrm{d}\varepsilon \\
&\le \inf_{\delta > 0} \left(\int_0^{\delta} \mathrm{d}\varepsilon + C_n \int_{\delta}^{+\infty} \frac{\mathrm{d}\varepsilon}{\varepsilon^{d/2}}\right) \\
&= \inf_{\delta > 0} \left(\delta + C_n \left(\frac{d}{2} - 1\right) \delta^{1-d/2}\right) \\
&= C_n^{2/d} \left(\frac{d}{2} - 1\right)^{4/d} \\
&= \frac{16R^2}{n^{2/d}} \left(\frac{d}{2} - 1\right)^{4/d},
\end{aligned}$$

which concludes the proof. □

Proof of Lemma 8.3 – Fix $k \in \{1,\dots,n\}$, $\rho > 0$ and $h \in \mathbf{H}_d(\rho)$ such that $\|h - h^*\|_\infty \leq \rho$. To ease notations we set

$$\hat{r}_h := \hat{r}_h[k] \quad \text{and} \quad W_i(h,.) := W_i[k](h,.), \quad i = n+1,\dots,2n.$$

First we have

$$\begin{aligned}
\mathbb{E}\Big[(r(X) - \hat{r}_h(h(X)))^2\Big] &\leq 2\mathbb{E}\Big[(r(X) - r_h(h(X)))^2\Big] \\
&\quad + 2\mathbb{E}\Big[(r_h(h(X)) - \hat{r}_h(h(X)))^2\Big].
\end{aligned}$$

Since $r = r_{h^*} \circ h^*$ and $r_{h^*} \in \mathscr{G}$, we have under assumptions (**A4**) and (**A5**)

$$\begin{aligned}
\mathbb{E}\Big[(r(X) - r_h(h(X)))^2\Big] &\leq 2\mathbb{E}\Big[(r_{h^*}(h^*(X)) - r_{h^*}(h(X)))^2\Big] \\
&\quad + 2\mathbb{E}\Big[(r_{h^*}(h(X)) - r_h(h(X)))^2\Big] \\
&\leq 2(L^2 + K^2)\|h - h^*\|_\infty^2 \\
&\leq 2(L^2 + K^2)\rho^2.
\end{aligned}$$

Therefore, we deduce that

$$\mathbb{E}\Big[(r(X) - \hat{r}_h(h(X)))^2\Big] \leq 4(L^2 + K^2)\rho^2 + 2\mathbb{E}\Big[(r_h(h(X)) - \hat{r}_h(h(X)))^2\Big].$$

Next, denoting $\mathscr{S} = \{X_{n+1},\dots,X_{2n}\}$, we have

$$\begin{aligned}
\mathbb{E}\Big[(r_h(h(X)) - \hat{r}_h(h(X)))^2 \Big| X\Big] &= \mathbb{E}\left[\left(r_h(h(X)) - \mathbb{E}\left[\hat{r}_h(h(X))\Big| X, \mathscr{S}\right]\right)^2 \Big| X\right] \\
&\quad + \mathbb{E}\left[\left(\hat{r}_h(h(X)) - \mathbb{E}\left[\hat{r}_h(h(X))\Big| X, \mathscr{S}\right]\right)^2 \Big| X\right] \\
&=: E_1 + E_2.
\end{aligned}$$

Here, we have used the fact that

$$\begin{aligned}
&\mathbb{E}\left[\left(r_h(h(X)) - \mathbb{E}\left[\hat{r}_h(h(X))\Big| X, \mathscr{S}\right]\right)\left(\hat{r}_h(h(X)) - \mathbb{E}\left[\hat{r}_h(h(X))\Big| X, \mathscr{S}\right]\right) \Big| X\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\left(r_h(h(X)) - \mathbb{E}\left[\hat{r}_h(h(X))\Big| X, \mathscr{S}\right]\right)\left(\hat{r}_h(h(X)) - \mathbb{E}\left[\hat{r}_h(h(X))\Big| X, \mathscr{S}\right]\right) \Big| X, \mathscr{S}\right] \Big| X\right] \\
&= \mathbb{E}\left[\left(r_h(h(X)) - \mathbb{E}\left[\hat{r}_h(h(X))\Big| X, \mathscr{S}\right]\right)\mathbb{E}\left[\hat{r}_h(h(X)) - \mathbb{E}\left[\hat{r}_h(h(X))\Big| X, \mathscr{S}\right] \Big| X, \mathscr{S}\right] \Big| X\right] \\
&= 0.
\end{aligned}$$

First, we bound the term E_1. We have

$$
\begin{aligned}
\mathbb{E}\left[\hat{r}_h(h(X))\Big|X,\mathscr{S}\right] &= \mathbb{E}\left[\sum_{i=n+1}^{2n} W_i(h,h(X))Y_i\Big|X,\mathscr{S}\right] \\
&= \sum_{i=n+1}^{2n} W_i(h,h(X))\mathbb{E}\left[Y_i|X_i\right] \\
&= \sum_{i=n+1}^{2n} W_i(h,h(X))r(X_i). \qquad (1.1)
\end{aligned}
$$

Therefore

$$
\begin{aligned}
E_1 &= \mathbb{E}\left[\left(r_h(h(X)) - \mathbb{E}\left[\hat{r}_h(h(X))\Big|X,\mathscr{S}\right]\right)^2\Big|X\right] \\
&= \mathbb{E}\left[\left(\sum_{i=n+1}^{2n} W_i(h,h(X))\left(r_h(h(X)) - r_{h^*}(h^*(X_i))\right)\right)^2\Big|X\right] \\
&\leq 3\mathbb{E}\left[\left(\sum_{i=n+1}^{2n} W_i(h,h(X))\left(r_h(h(X)) - r_{h^*}(h(X))\right)\right)^2\Big|X\right] \\
&\quad + 3\mathbb{E}\left[\left(\sum_{i=n+1}^{2n} W_i(h,h(X))\left(r_{h^*}(h(X)) - r_{h^*}(h(X_i))\right)\right)^2\Big|X\right] \\
&\quad + 3\mathbb{E}\left[\left(\sum_{i=n+1}^{2n} W_i(h,h(X))\left(r_{h^*}(h(X_i)) - r_{h^*}(h^*(X_i))\right)\right)^2\Big|X\right] \\
&=: J_1 + J_2 + J_3.
\end{aligned}
$$

Since $r = r_{h^*} \circ h^*$ and $r_{h^*} \in \mathscr{G}$, assumptions (**A4**) and (**A5**) lead to

$$J_1 \leq 3K^2\rho^2 \quad \text{and} \quad J_3 \leq 3L^2\rho^2.$$

The fact that r_{h^*} is L-Lipschitz leads to

$$J_2 \leq 3L^2\mathbb{E}\left[\left(\frac{1}{k}\sum_{i=1}^{k} \|X^h - X^h_{(i)}(X^h)\|\right)^2\Big|X\right].$$

Now let $\tilde{n} = \lfloor n/k \rfloor$. We split the sample $\left\{X^h_{n+1}, \ldots, X^h_{n+k\tilde{n}}\right\}$ into k subsamples of size $\tilde{n}$:

$$Z_j = \left\{X^h_{n+(j-1)\tilde{n}+1}, \ldots, X^h_{n+j\tilde{n}}\right\}, \quad j = 1, \ldots, k,$$

and denote by $Z_j^{(1)}$ the closest element of Z_j from X^h. Then

$$\sum_{j=1}^{k} \|X^h_{(j)}(X^h) - X^h\| \leq \sum_{j=1}^{k} \|Z_j^{(1)} - X^h\|.$$

Therefore, by Jensen's inequality and lemma 1.1, we have

$$\begin{aligned}
\mathbb{E}\left[E_1\right] &\leq \frac{3L^2}{k} \sum_{j=1}^{k} \mathbb{E}\|Z_j^{(1)} - X^h\|^2 + 3(L^2+K^2)\rho^2 \\
&\leq \frac{48L^2R^2}{\tilde{n}^{2/d}} \left(\frac{d-2}{2}\right)^{4/d} + 3(L^2+K^2)\rho^2 \\
&\leq 48L^2R^2 4^{1/d} \left(\frac{d-2}{2}\right)^{4/d} \left(\frac{k}{n}\right)^{2/d} + 3(L^2+K^2)\rho^2,
\end{aligned}$$

where the last inequality holds provided $\frac{k}{n}\lfloor n/k \rfloor \geq \frac{1}{2}$.

Now we turn to bounding the term E_2. According to (1.1), we have

$$\begin{aligned}
E_2 &= \mathbb{E}\left[\left(\sum_{i=n+1}^{2n} W_i(h, h(X))(Y_i - r(X_i))\right)^2 \middle| X\right] \\
&= \mathbb{E}\left[\sum_{i=n+1}^{2n} W_i(h, h(X))^2 (Y_i - r(X_i))^2 \middle| X\right],
\end{aligned}$$

where we have used the fact that if $i \neq j \in \{n+1, \ldots, 2n\}$

$$\mathbb{E}\left[W_i(h, h(X))(Y_i - r(X_i)) W_j(h, h(X))(Y_j - r(X_j)) \middle| X\right] = 0.$$

Using the properties $\sum_{i=n+1}^{2n} W_i(h, h(X)) = 1$, $W_i(h, h(X)) \leq \frac{1}{k}$ and $|Y_i - r(X_i)| \leq B + L$, for all $i \in \{n+1, \ldots, 2n\}$, we conclude that

$$\mathbb{E}\left[E_2\right] \leq \frac{(B+L)^2}{k}.$$

To complete the proof, write

$$\begin{aligned}
\mathbb{E}\left[(r(X)-\hat{r}_h(h(X)))^2\right] &\leq 2\mathbb{E}[E_1]+2\mathbb{E}[E_2]+4(L^2+K^2)\rho^2 \\
&\leq \frac{2(B+L)^2}{k}+96L^2R^2 4^{1/d}\left(\frac{d-2}{2}\right)^{4/d}\left(\frac{k}{n}\right)^{2/d}+10(L^2+K^2)\rho^2 \\
&\leq C\left\{\frac{1}{k}+\left(\frac{k}{n}\right)^{2/d}\right\}+C\rho^2,
\end{aligned}$$

where $C := \max\left\{2(B+L)^2;\ 10(L^2+K^2)\ ;\ 96L^2R^2 4^{1/d}\left(\frac{d-2}{2}\right)^{4/d}\right\}$. $\square$

References

Cucker, F. and Smale, S. (2001). On the mathematical foundations of learning theory. *Bulletin of the American Mathematical Society*. Vol. 39, pp. 1-49.